IBM Developer
SKILLS NETWORK

# Winning Space Race with Data Science


Kailash Pushpa Rajamanickam

Feb 03, 2023

# Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

# Executive Summary

**Methodologies**

- Data collection using API Calls & Web scraping
- Data wrangling to understand the data better
- Exploratory data analysis (EDA) using visualization and SQL
- Interactive visual analytics using Folium and Plotly Dash
- Predictive analysis using Logistic Regression, Support vector method, Decision Tree and K nearest neighbour models
- Evaluation of models using accuracy score and confusion matrix

**Results**

- The success or failure of landing the first stage is influenced by the Payload Mass, Orbit and Launch Site.
- Interactive analytics demo in screenshots
- Predictive analysis using Decision Tree model is the best choice for this project

# Introduction

## Background

The cost of launching a SpaceX Falcon 9 rocket is $62M. This is much lesser than other providers, who charge $165M or more.

The huge price difference is because only SpaceX has the technology to reuse the first stage of their rockets.

## Problem

Reuse is possible only if the first stage comes back to earth and lands successfully without any damage. SpaceX has still not managed to achieve 100% success in safely landing the first stage.

This project aims to predict the success of landing the first stage of SpaceX Falcon 9 rocket. This will help the competitor SpaceY in predicting the cost of launch before bidding against SpaceX.

Section 1

# Methodology

# Methodology

The following processes and tools/techniques were used in this project:

- Data collected from
  - SpaceX REST API
  - Wikipedia (web scraping using BeautifulSoup library)
- Data wrangling, which includes
  - Removal of unwanted data from data set
  - Replacing invalid data (empty, null, NaN, etc) with appropriate values
- Exploratory data analysis (EDA) using
  - Visualization (matplotlib)
  - SQL

# Methodology (contd.)

- Interactive visual analytics using
  - Folium (geographic analysis of launch sites)
  - Plotly Dash (interactive dashboard to analyze correlation between parameters)
- Predictive analysis using classification models
  - Logistic Regression
  - Square Vector Method
  - Decision Tree
  - K Nearest Neighbor
- Model evaluation using Accuracy Score and Confusion Matrix

# Data Collection

This project requires information about:

- SpaceX rockets
- Orbits
- Launch sites
- Historical data about the success/failure of launches
- and so on …

The data is collected from two sources:

- **SpaceX API**
  https://api.spacexdata.com/v4/
- **Wiki page: List of Falcon and Falcon 9 Heavy Launches**
  https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

# Data Collection – SpaceX API

The data collected from SpaceX API is cleaned, filtered, parsed, and converted into Pandas data frame for easy analysis.

The flowchart on the right explains the steps followed.

Get data of all SpaceX launches

⇩

Filter and get 'Falcon 9' launches only

⇩

Remove unwanted parameters from data set

⇩

Identify records with missing data

⇩

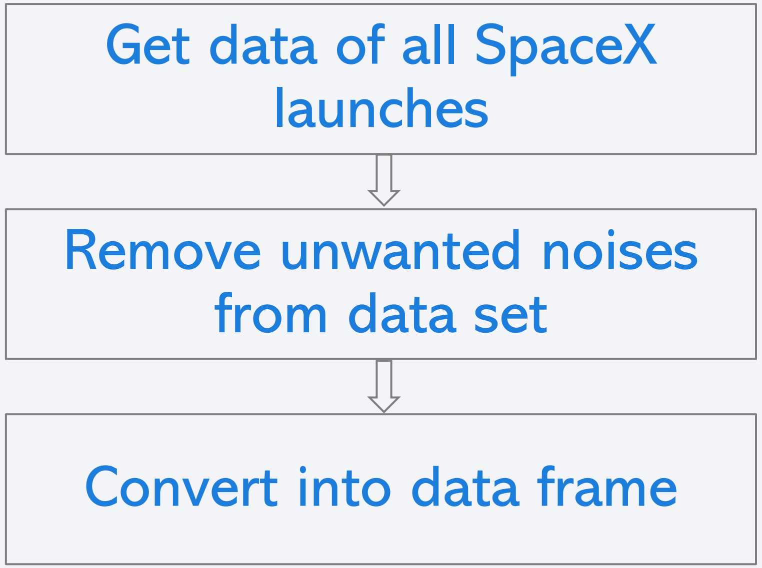
Handle missing info using appropriate technique

https://github.com/kailashpr/capstone/blob/main/M1L1-Data%20Collection.ipynb

# Data Collection – Scraping

The data collected from the wiki page is cleaned, filtered, parsed, and converted into Pandas data frame for easy analysis.

The flowchart on the right explains the steps followed.




https://github.com/kailashpr/capstone/blob/main/M1-L2-Web%20Scraping.ipynb

# Data Wrangling

**Metrics calculated:**

- Number of launches on each site
- Number of launches for each orbit
- Number of successes for each orbit type
- Overall success rate – 0.66666666666

https://github.com/kailashpr/capstone/blob/main/M1L3-Data%20Wrangling.ipynb

# EDA with Data Visualization

- Success is more frequent when Flight number > 60.
- Launches from KSC LC 39A have succeeded more.
- The heavier the payload, the more the probability of success.
- Launches to orbits ES-L1, GEO, HEO, and SSO have seen more success.
- The frequency of success has increased considerably in the last few years.
- No correlation seen between Flight Number and Orbit Type
- No correlation seen between Payload and Orbit Type

https://github.com/kailashpr/capstone/blob/a175f340dee66be09610987d231768308112673c/M2-Data%20Preparation.ipynb

# EDA with SQL

**Data analyzed using SQL**

- List of launch sites
- Total payload weight sent by NASA alone
- Average payload mass of SpaceX Falcon 9 rockets
- Date of first successful landing on ground pad
- Number of failed & successful missions
- List of boosters that carried the max payload mass
- Boosters with 4000 to 6000 kgs of weight that have successfully landed on drone ship
- Failed landing outcomes in drone ship, their booster versions, and launch site names in year 2015
- Rank landing outcomes between 2010-06-04 and 2017-03-20

https://github.com/kailashpr/capstone/blob/main/M2L2-EDA%20SQL.ipynb

# Build an Interactive Map with Folium

- All launch sites of SpaceX
  are grouped in just
  two geographic clusters
- Some launch sites have seen more success than the others
- No correlation is seen between the landing outcome & the geographic cluster to which the Launch Site belongs

https://github.com/kailashpr/capstone/blob/a175f340dee66be09610987d231768308112673c/M3-Location%20Analysis%20with%20Folium.ipynb

# Build a Dashboard with Plotly Dash

Interactive Dashboard helped in identify how different parameters of the launch correlate to the successful landing of the first stage:

| *Parameter* | *Preferred Option* | *Reason* |
|---|---|---|
| Launch Site | KSC-LC-39A | Most number of successful launches (7 launches)<br>Highest success rate (76.9%) |
| Payload Weight | 2-3 tonnes | Most number of successes (against the range 6-8 tonnes which has the failures) |
| Booster Version | FT | Highest success rate among all versions<br>(15 successes in 23 attempts) |

https://github.com/kailashpr/capstone/blob/b0789c84cf097d78321c2ce73958006a2ed652b1/spaces_dash_app.py

# Predictive Analysis (Classification)

The sample data set was split into training and test sets, and were evaluated on 4 different machine learning models:

| *Model* | *Accuracy Score* | *Confusion Matrix Result* | *Conclusion* |
|---|---|---|---|
| Logistic Regression | 0.83333 | No | Rejected |
| Support Vector Machine | 0.8333 | No | Rejected |
| Decision Tree | 0.7777 | **Yes** | **Selected** |
| K Nearest Neighbor | 0.8333 | No | Rejected |

https://github.com/kailashpr/capstone/blob/main/M4-Machine%20Learning%20Prediction.ipynb

# Results

- The success or failure of landing the first stage is influenced by the Payload Mass, Orbit and Launch Site.
- Interactive analytics demo in screenshots
- Predictive analysis using Decision Tree model is the best choice for this project

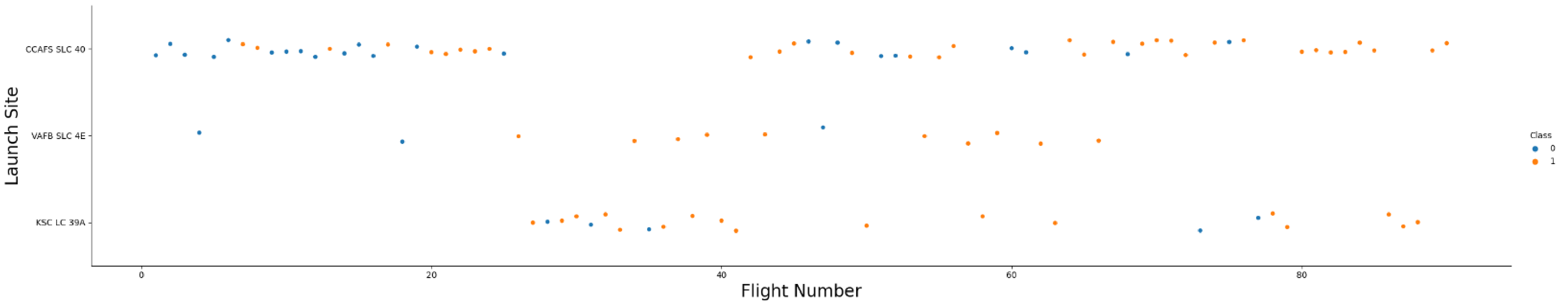
Section 2

# Insights drawn from EDA

# Flight Number vs. Launch Site

Success is more frequent when

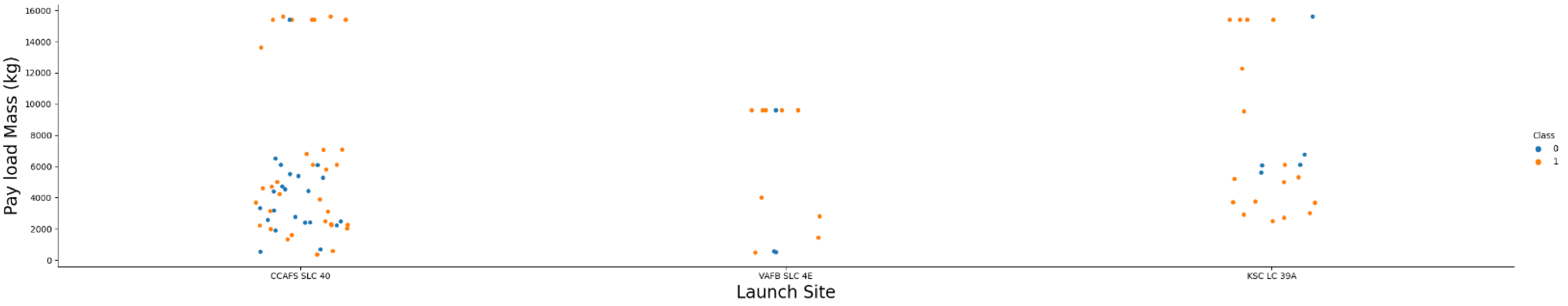
- Flight number > 60
- Launched from site - KSC LC 39A



# Payload vs. Launch Site

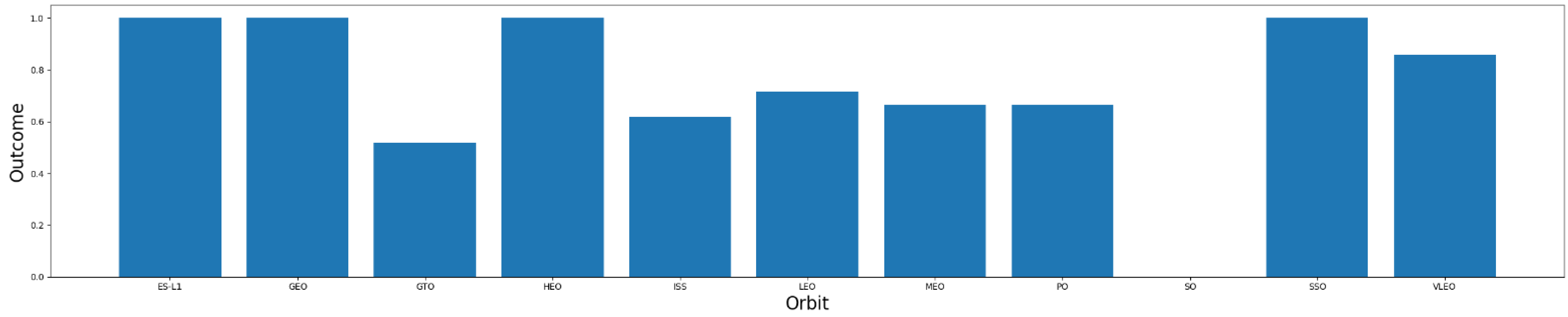
Launches from KSC LC 39A are more likely to succeed.

Also, the heavier the payload, the more the probability of success.



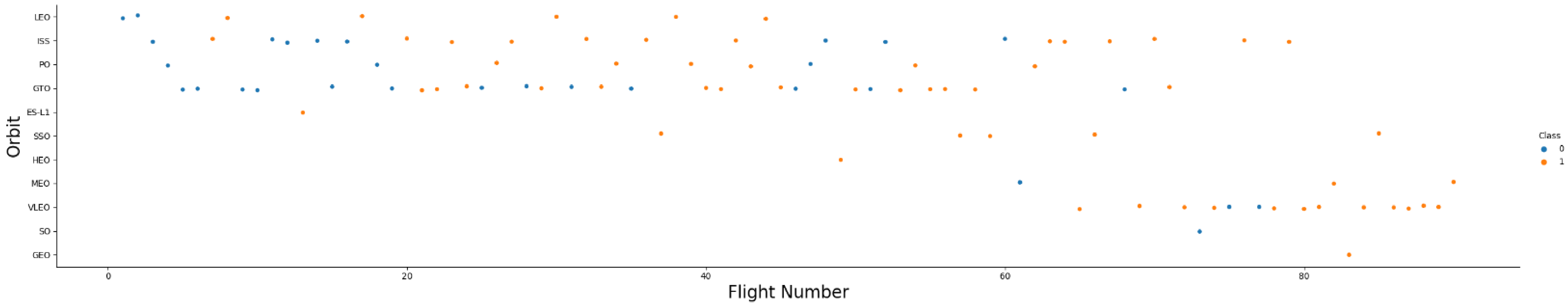
# Success Rate vs. Orbit Type

Launches to orbits ES-L1, GEO, HEO, and SSO have seen more success



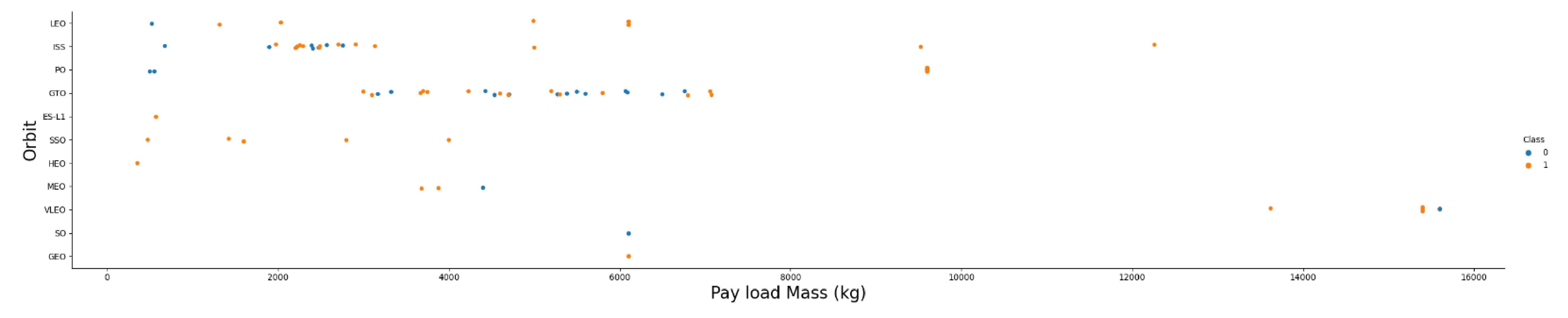
# Flight Number vs. Orbit Type

No correlation is seen between Flight Number and Orbit Type



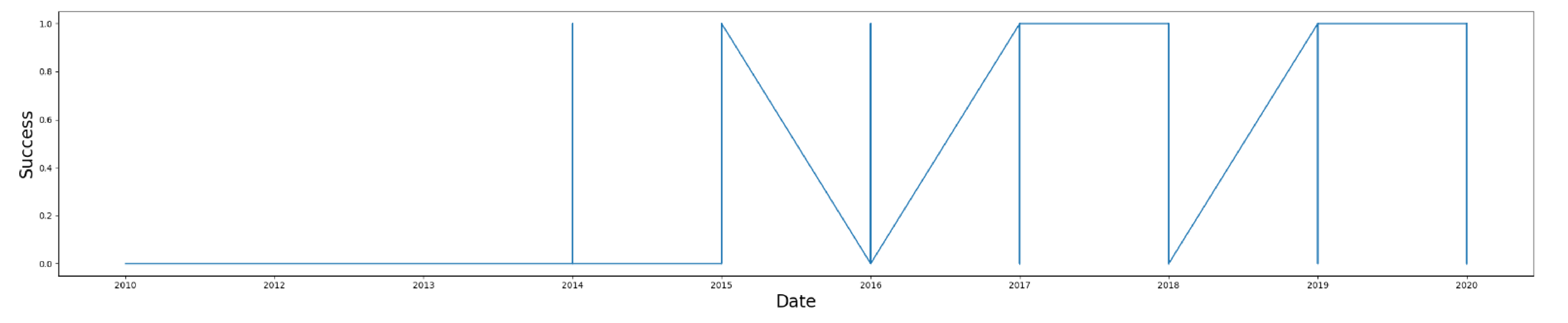
# Payload vs. Orbit Type

No correlation is seen between Payload and Orbit Type



# Launch Success Yearly Trend

The frequency of success has increased considerably in the last few years.



# All Launch Site Names

There are 4 launch sites:

| Launch_Site |
|---|
| CCAFS LC-40 |
| VAFB SLC-4E |
| KSC LC-39A |
| CCAFS SLC-40 |

# Launch Site Names Begin with 'CCA'

Details of a few launches where the name of the launch site begins with “CCA”:

| Date | Time (UTC) | Booster_Version | Launch_Site | Payload | PAYLOAD_MASS__KG_ | Orbit | Customer | Mission_Outcome | Landing _Outcome |
|---|---|---|---|---|---|---|---|---|---|
| 04-06-2010 | 18:45:00 | F9 v1.0 B0003 | CCAFS LC-40 | Dragon Spacecraft Qualification Unit | 0 | LEO | SpaceX | Success | Failure (parachute) |
| 08-12-2010 | 15:43:00 | F9 v1.0 B0004 | CCAFS LC-40 | Dragon demo flight C1, two CubeSats, barrel of Brouere cheese | 0 | LEO (ISS) | NASA (COTS) NRO | Success | Failure (parachute) |
| 22-05-2012 | 07:44:00 | F9 v1.0 B0005 | CCAFS LC-40 | Dragon demo flight C2 | 525 | LEO (ISS) | NASA (COTS) | Success | No attempt |
| 08-10-2012 | 00:35:00 | F9 v1.0 B0006 | CCAFS LC-40 | SpaceX CRS-1 | 500 | LEO (ISS) | NASA (CRS) | Success | No attempt |
| 01-03-2013 | 15:10:00 | F9 v1.0 B0007 | CCAFS LC-40 | SpaceX CRS-2 | 677 | LEO (ISS) | NASA (CRS) | Success | No attempt |

# Total Payload Mass

Total Payload Mass carried by boosters from NASA:

| SUM(PAYLOAD_MASS__KG_) |
| ---: |
| 45596 |

# Average Payload Mass by F9 v1.1

Average payload mass carried by booster version F9 v1.1

| AVG(PAYLOAD_MASS__KG_) |
| ---: |
| 2534.6666666666665 |

# First Successful Ground Landing Date

The first successful landing on ground pad was achieved on

| MIN(DATE) |
| --- |
| 01-05-2017 |

# Successful Drone Ship Landing with Payload between 4000 and 6000

Boosters that have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000:

| Date | Time (UTC) | Booster_Version | Launch_Site | Payload | PAYLOAD_MASS__KG_ | Orbit | Customer | Mission_Outcome | Landing _Outcome |
|---|---|---|---|---|---|---|---|---|---|
| 06-05-2016 | 05:21:00 | F9 FT B1022 | CCAFS LC-40 | JCSAT-14 | 4696 | GTO | SKY Perfect JSAT Group | Success | Success (drone ship) |
| 14-08-2016 | 05:26:00 | F9 FT B1026 | CCAFS LC-40 | JCSAT-16 | 4600 | GTO | SKY Perfect JSAT Group | Success | Success (drone ship) |
| 30-03-2017 | 22:27:00 | F9 FT B1021.2 | KSC LC-39A | SES-10 | 5300 | GTO | SES | Success | Success (drone ship) |
| 11-10-2017 | 22:53:00 | F9 FT B1031.2 | KSC LC-39A | SES-11 / EchoStar 105 | 5200 | GTO | SES EchoStar | Success | Success (drone ship) |

# Total Number of Successful and Failure Mission Outcomes

Total number of successful and failure mission outcomes:

| A | COUNT(*) |
|---|---|
| Failure | 1 |
| Success | 100 |

# Boosters Carried Maximum Payload

Boosters that have carried the maximum payload mass

| Booster_Version | |
|---|---|
| F9 B5 B1048.4 | F9 B5 B1049.5 |
| F9 B5 B1049.4 | F9 B5 B1060.2 |
| F9 B5 B1051.3 | F9 B5 B1058.3 |
| F9 B5 B1056.4 | F9 B5 B1051.6 |
| F9 B5 B1048.5 | F9 B5 B1060.3 |
| F9 B5 B1051.4 | F9 B5 B1049.7 |

# 2015 Launch Records

Failed landing outcomes in drone ship, their booster versions, and launch site names in the year 2015

| YEAR | MONTH | COUNT(*) |
|---|---|---|
| 2015 | 01 | 1 |
| 2015 | 04 | 1 |

# Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

Ranking of the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order:

| YEAR | SUCCESS_COUNT |
|---|---|
| 2016 | 5 |
| 2017 | 3 |
| 2015 | 1 |

Section 3

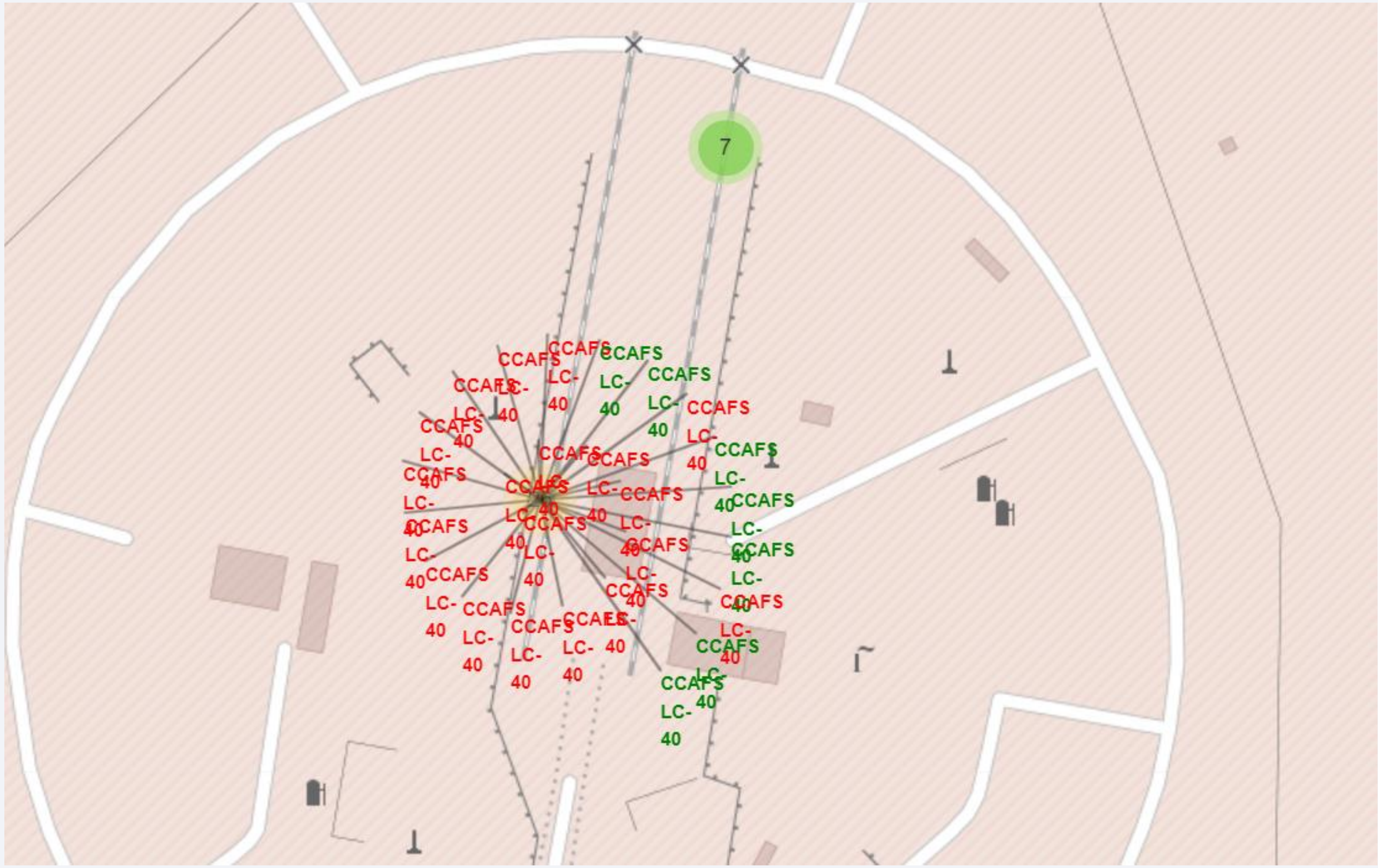
# Launch Sites Proximities Analysis

# Launch Sites - Clusters


All launch sites of SpaceX are grouped in just **two** geographic clusters

# Launch Sites vs Mission Outcome (Success / Failure)

Some launch sites have seen more success than the others



# Proximity between Launch Sites

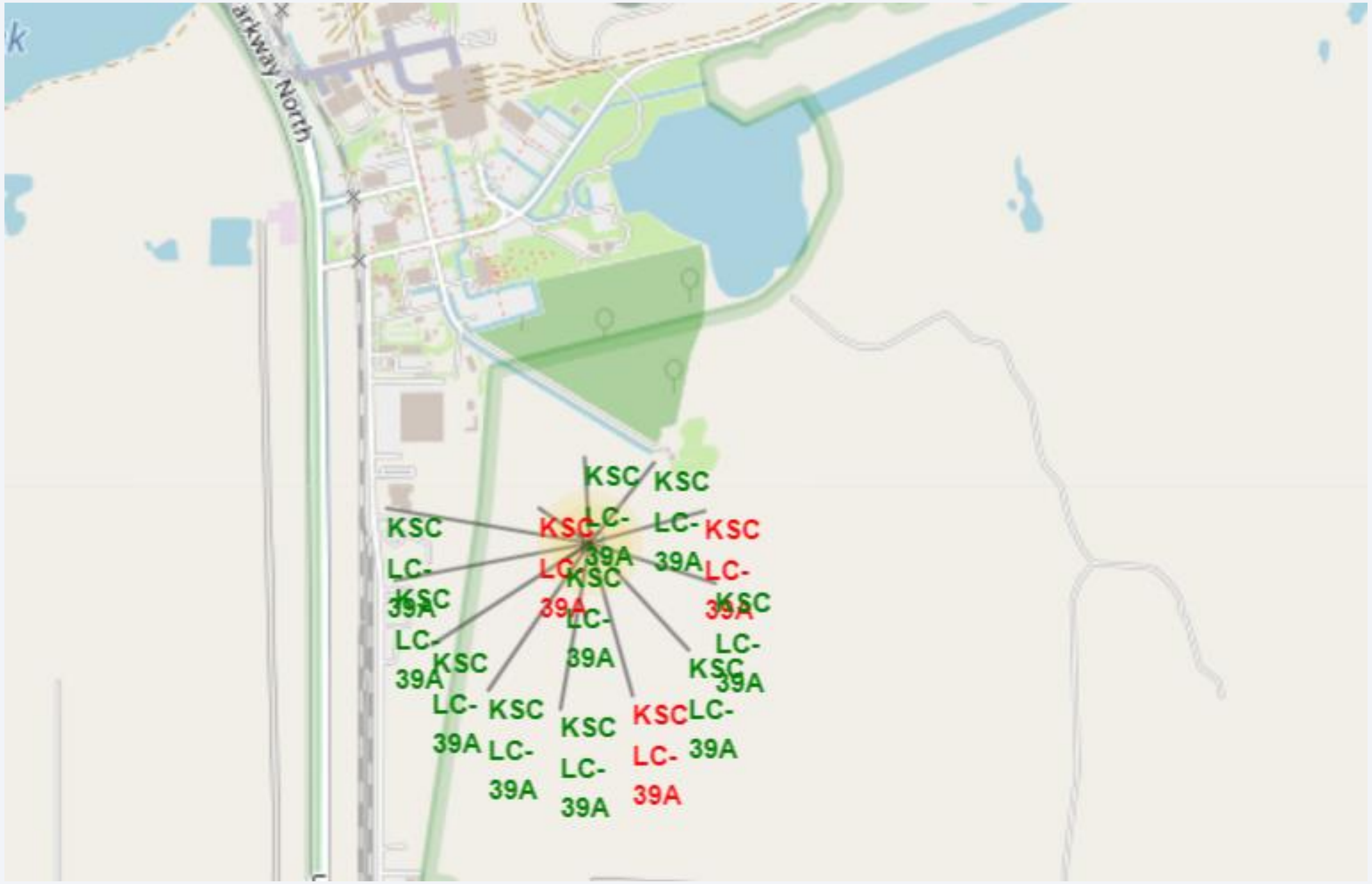
No correlation is seen between

the landing outcome

&

the geographic cluster

to which the Launch Site belongs

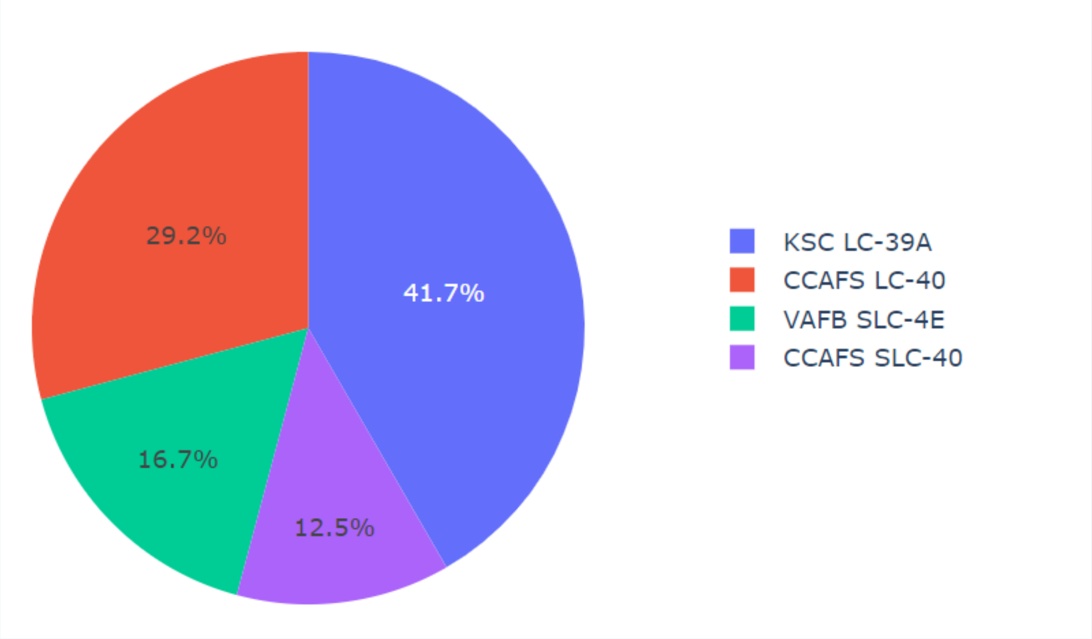


Section 4

# Build a Dashboard with Plotly Dash

# Success Count by Launch Site

**41.7%** of all successful missions were launched from **KSC LC-39A**

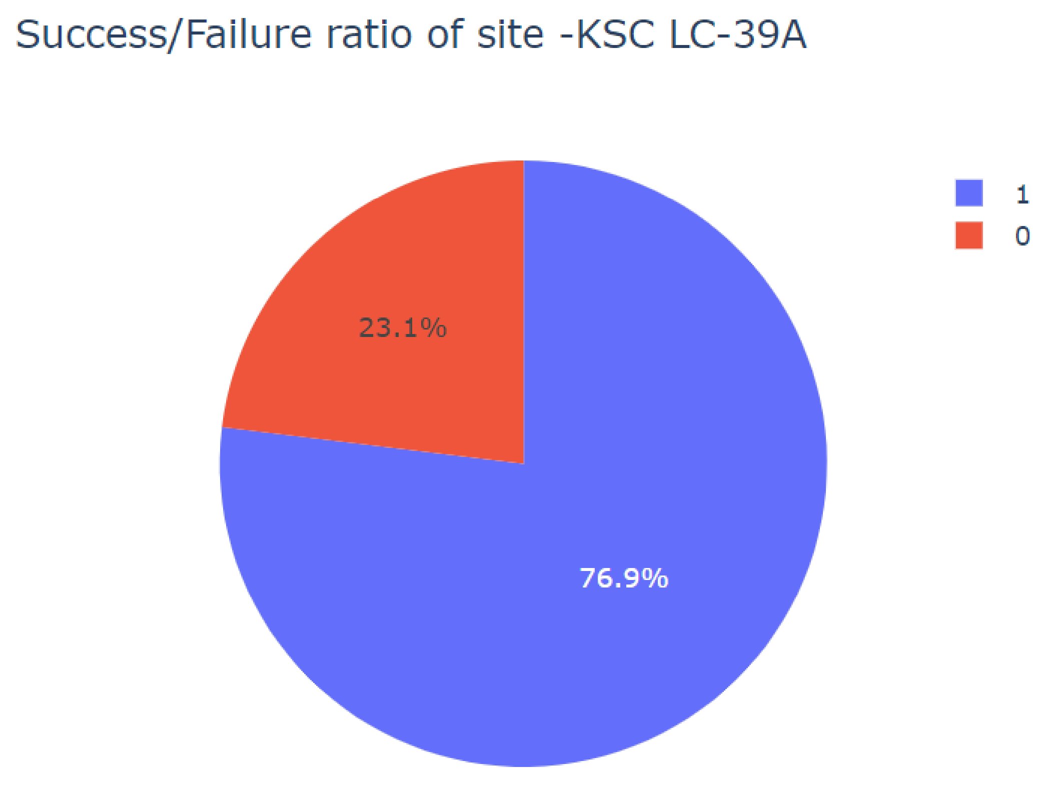


# Launch Site with Highest Success Rate

Launch with the highest success rate is

KSC LC-39A

- 76.9% of the missions launched from this site have succeeded.
- Only 23.1% have failed.



# Payload vs Success Rate

- Launch site KSC-LC-39A has the most number of successes
- Payloads in the range of 2-3 tonnes had the most number of sucesses (against the 6-8 tonnes range that has failed most).
- Booster version FT has the highest success rate  among all versions (15 successes & 8 failures)

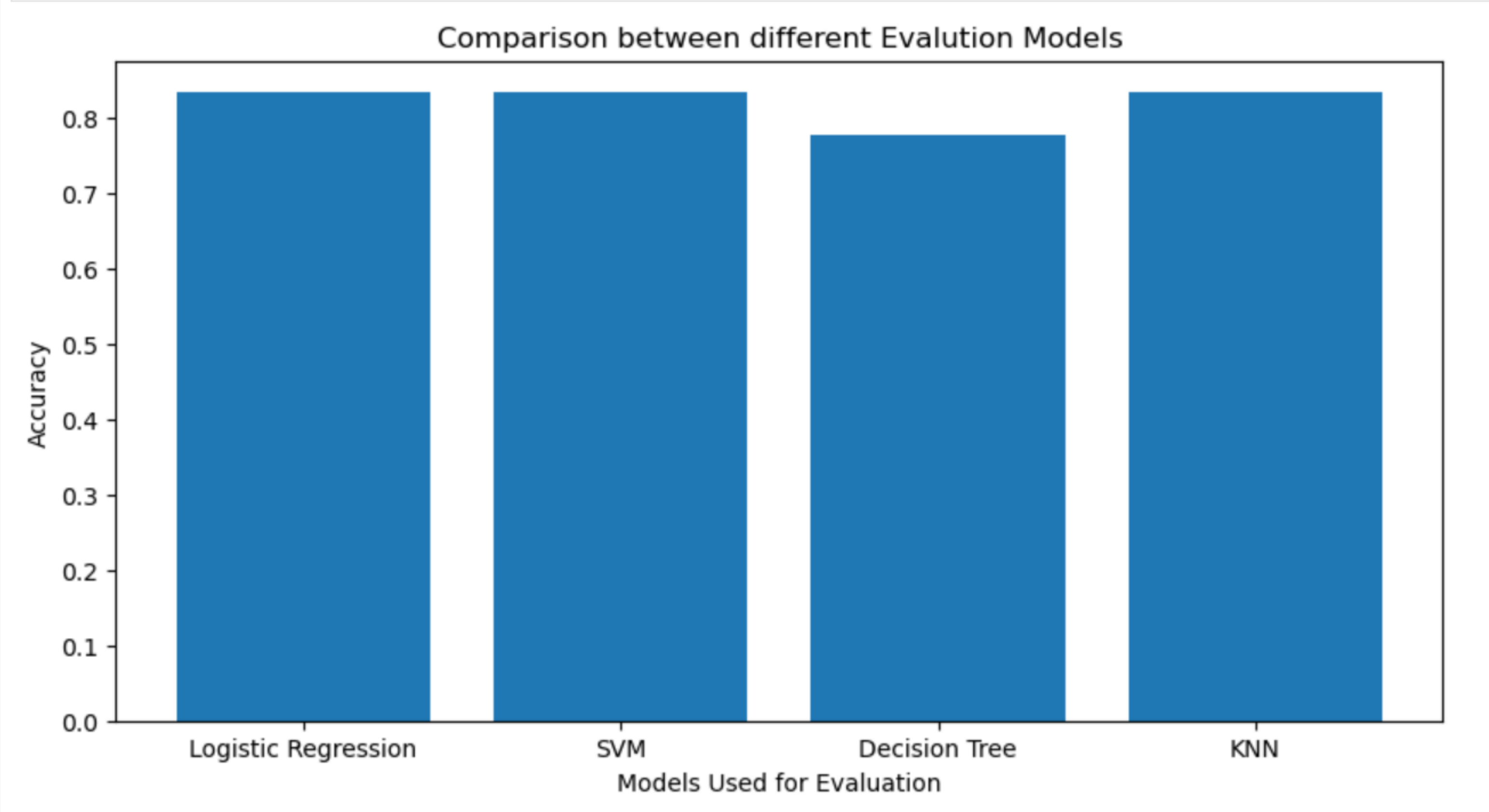
Section 5

# Predictive Analysis (Classification)

# Classification Accuracy

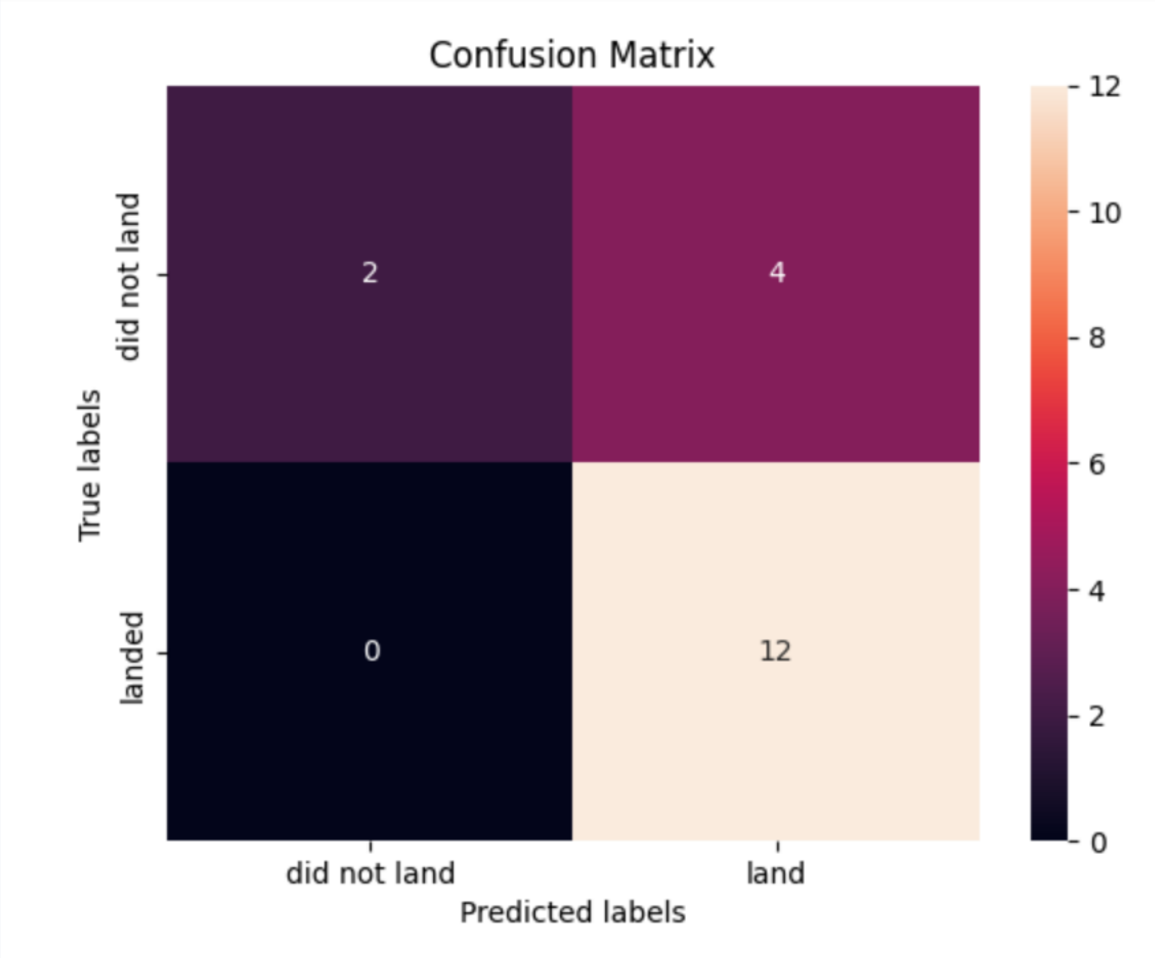
Of the 4 models used for evaluation, the accuracy was highest (0.833) for 3 models - Logistic regression, Support vector machine, and K nearest neighbor methods

The Decision Tree model had slightly lower (0.777) accuracy.



# Confusion Matrix

The Decision Tree model performed the best when evaluated using confusion matrix.



# Conclusions

It is concluded that the **Decision Tree** model is the best model to predict whether the first stage of the SpaceZ Falcon 9 rocket will land successfully.

The success or failure is influenced by the following parameters:

- Payload Mass
- Orbit
- Launch Site

# Appendix

Python Notebooks used in this project are archived at:

https://github.com/kailashpr/capstone

Thank you!